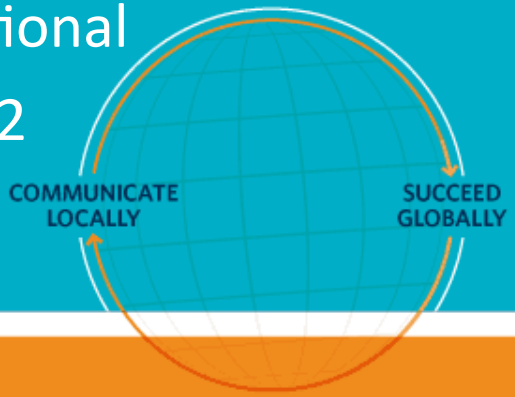
Project Ideas
for LT-Innovate workshop, Oct 8

Serge Gladkoff, GALA Standards Director
President, Logrus International

Brussels, October 2012



Current [dizzzoo]array

- 100 types of DATA
- 1100 CMS systems
- 50 TMS systems
- 10 TM tools
- 10 major MTs and 2000 Moses systems ☺
- 50 LANGUAGES

= 27,500,000,000 problems

What interoperability?

Current disconnect:

- Process friction is high and only seems to be getting worse;
- MT is not ready for deployment, requires technological level beyond LSP capabilities;
- Technologies are like zoos; standards are not implemented;
- Clients, however, demand price level as if HQMT is real;
- LSPs are forced to juggle with skills and tools that have no relation to translation, language and SME knowledge;
- No unified technological platform exists;
- Technical overhead is too high; skilled translators and PMs spend time on operations which are not paid for by the clients;
- Technologies in general are either too expensive or proprietary;
- Tools developers seem to be absorbed with their own vision;
- Data (content) creation process lacks understanding of localization needs;
- Lack of rich metadata to ease multilingual translation process.

What next?

- We need removal of duplication of effort that will engage the most advanced ideas of today (cloud);
- With so many various technologies clearly there's a big and concrete need for innovative language technologies;
- We need serious open investment of effort into the translation technology of the future and collaboration;
- Standardization and interoperability are key;
- Data is the key; we need to change the authoring process; add massive metadata; standardize; create common, non-proprietary technology platform for productive engagement.

Reference Slide for Ideas 1, 2 and 3. Killer Development Method: Open Source based on Open Standards

- Almost all the "finished" technologies are proprietary.
- EU will have issues around copyright, trademark, patents, etc.
- Proprietary technologies and tools will inhibit innovation, information and technology exchange, will tie small suppliers into closed camps".
- With so many various technologies clearly there's a big and concrete need for innovative language technologies, but these technologies must be open source so that develop the framework and platform for the larger industry involvement, to gauge serious open investment of effort into the translation technology of the future and collaboration from SMALL players;
- The goal is to create common, non-proprietary technology platform for productive engagement.
- Finally, the language industry is small and currently there's an engagement barrier of LSPs into the research and development for the future.

Reference Slide for Ideas 2 and 3. Killer Weapon: Open Standards Killer Standard: HTML5

The success of Apple, the company that managed to brings it products nearly to every user of the world, is based on relying on open standards. There's one standard that is used by all – HTML5.

"Apple has adopted HTML5, CSS and JavaScript – all open standards. Apple's mobile devices all ship with high performance, low power implementations of these open standards. HTML5, the new web standard that has been adopted by Apple, Google and many others, lets web developers create advanced graphics, typography, animations and transitions without relying on third party browser plug-ins (like Flash). HTML5 is completely open and controlled by a standards committee."

There are open standards that rely on HTML5, such as WebKit, a complete open-source HTML5 rendering engine that is the heart of the Safari web browser used in all our products. WebKit has been widely adopted. Google uses it for Android's browser, Palm uses it, Nokia uses it. Microsoft is adopting HTML5 in its new Office 2013 application.

"New open standards created in the mobile era, such as HTML5, will win on mobile devices (and PCs too). Perhaps Adobe should focus more on creating great HTML5 tools for the future, and less on criticizing Apple for leaving the past behind." – perhaps we should focus on HTML5 for the future effort for multilingual Europe, rather than on proprietary technologies?

Thoughts on Flash, Steve Jobs

http://www.apple.com/hotnews/thoughts-on-flash/

Actual industry in not engaged into the future development; GALA Europe is ready to foster

- 170 European companies
- 55% of GALA membership
- 32 European countries
- Most active companies
- LSPs, technology, clients
- They all use INTERNET ☺

Idea 1: An EU LT Incubator

(ICT-2013.10.3, ICT-2013.11.1, ICT-2013.11.4 ???)

Situation.

EU has funded a lot of projects in the past. Many of the great results are open for the world, and the world is taking the advantage. Significant portion of EU funding has been actually used to promote the competitiveness of other regions (examples: Google Translate, Siri are made with the results of EU funding), rather than promote the competitiveness of the EU which has funded this innovative research. EU has funded over $150 million Euros in 60 research institutions, but many of these projects were not surviving and did not result into any self-sustainable business. Commercial companies are using the EU-funded research to take to their local headquarters and boost their competitiveness. Large commercial companies use their R&D department to create the missing link between the results of the research and commercial, proprietary projects.

Solution.

Establish a collective "R&D implementation incubator" for European LT, with the goals:

- To catalogue the EU-funded research already done and funded by EU, along with the results.
- To create the platform for ideas incubator; engage GALA European members
- To create an open platform for R&D based on such research to enable European LSPs to implement these products and deliver them to the users and Europe, creating value for Europe.
- To create open tools based on open standards in quest for the goals of the Calls.

Implementation

- Add relevant functionality to existing GALA projects and systems, Tools Corner and GALA Connect.

Idea 2: (ICT-2013.9.8 , ICT-2013.11.3 ???)
Editing in the HTML5 browser

Theme 3: Using the browser instead of specialized CAT tools to translate HTML5 pages

Problem description:

Various, often incompatible, proprietary and expensive CAT tools are often required to translate content. The requirements to technical qualification of translators are getting too high, which is especially problematic for emerging languages and community projects. The list of "qualified" translators for that reason is getting too narrow to satisfy current growing demands, productivity is suffering and the publication process is long.

Proposed solution of this problem:

The content can be translated in the HTML5 format on the client side directly in browser, with the return of translations to the database server and immediate publication for preview over the course of the translation process. Such technology can be used to share and split content translation in an online mode between many project participants. This can support crowdsourcing, community translation and machine translation methods and technologies. Publication cycle can be shortened.

Secondary problem:

Research is required to implement bi-directional data exchange between the server and browser in the course of the translation process directly in the browser. Code has to be developed (in the form of scripts or add-ons) to implement the browser interface (windows) and data exchange with the server.

Project result:

Description of the content translation in HTML5 format on the client side directly in the browser with the pushback of data to the server modeled on the actual client-server system

Sample HTML5 page and code (add-on or page script) to translate content directly in the browser modeled on actual client-server system

Description of the translation process with screenshots to test given content samples and code (generic example)

Idea 3: (for ICT-2013.4.1 ???)
Tailor HTML5 ITS for EU

Background

Nowadays, content hosting and rendering technology transitions quickly from closed client-server systems to cloud-based web applications communicating through open standards, such as HTML5 and XLIFF, be it on PC-based or mobile devices. The data of cloud-based, mobile applications will inevitably be wrapped into tagged standards and metadata. Compliance with open standards will decrease friction and improve interoperability. As applications and systems take off into the cloud and mobile applications, they will also be wrapped in standard tagged formats. The data will travel between systems and systems and systems and humans for interactions, data transfer, control, delivery, perception; the content will be translated. It is therefore quite necessary to have context information embedded in data as opposed to metadata, and the way metadata is tagged should be standardized.

For example, standard EU information tags may be passed via web, tagged by certain differentiators. If they are tagged in a standard fashion, system communication is guaranteed to be run smoothly, without incompatibilities. Also, the standard way of managing user elements will help in the translation stage.

Variety on content to be translated. Variety of technologies and platforms people work with. Delivery of EU content to all citizen.

Problem Description

Localization is handled in proprietary systems. These systems require lots of knowledge and additional information embedded in the form of additional instructions, systems or both. The ease of tools will facilitate knowledge accumulation and retention. The formats and data containers are important. Standardizing data formats and containers is key to solution of these problems.

Proposed Solution

To tailor ITS for EU metadata in order for future standardization of tools. For example, standard “middleware” data, such as quality dimensions, for example; or standard EU-related reference data, etc. This will enable future rendering of EU-related information in any browser, on any platform, for any process, including copyediting and translation.

Thank you!

www.gala-global.org

standards@gala-global.org

sgladkoff@gala-global.org

About GALA

The Globalization and Localization Association is the largest global non-profit association within the language industry, providing resources, education, ideas and research for companies working with translation services, language technology and content localization. GALA member companies are vendors and buyers of language services and technologies. They deploy sophisticated multilingual strategies and proven tools to take content and products to markets around the world.